DOCUMENT RESUME

ED 241 930 CS 208 072

AUTHOR Marzano, Robert J.
TITLE Toward a Working Model for the Analysis of Cohesion and Coherence in Writing.
PUB DATE 83
NOTE 33p.
PUB TYPE Viewpoints (120) -- Information Analyses (070)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Coherence; *Cohesion (Written Composition); Communication Research; *Connected Discourse; *Discourse Analysis; *Models; Research Design; Research Methodology; Systems Analysis; *Systems Approach
IDENTIFIERS *Text Structure

ABSTRACT

Although most models of connected discourse are strikingly similar in the types of relationships they describe, they are strikingly different in two areas: the unit of analysis and the dimensions on which cohesion versus coherence are described. Common systems for analyzing written text use the sentence, T-unit, clause, or proposition as the unit of analysis. However, these units of analysis are often confounded with the issue of relationship types. That is, when the unit of analysis is large, salient relationships between unit elements are lost and when the unit of analysis is small, too many relationships are identified. A compromise unit of analysis might be predication, a system that recognizes both finite and nonfinite clauses and contains two types of relationships--argumental and nonargumental. Since the identification of referential ties (cohesion) is one of the first lines of discourse processing, it is logical to use a model that identifies lexical references from one predication to the next. Coherence can be determined by applying a set of macrorules to the discourse as a whole. For a set of predications, the macrostructure may be signaled by predications summarizing or subsuming the other predications in the set. There are also several types of patterns of relationships or macropatterns, each with governing predications. The advantage of this system of analysis is that it permits for a quantitative analysis of discourse data. (HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED241930

TOWARD A WORKING MODEL FOR THE ANALYSIS OF COHESION AND COHERENCE IN WRITING

by

Robert J. Marzano, Ph.D.

Mid-continent Regional Educational Laboratory

Summer, 1983

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Robert J. Marzano

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS 208 072

Currently there are a number of systems or models with which a writing researcher may describe and quantify the ways various units of thought are related within connected discourse. Although most models are strikingly similar in the types of relationships they describe, they are strikingly different in two areas: 1) the unit of analysis and 2) the dimensions on which cohesion vs. coherence are described. In this paper I will review how some of the more commonly used systems address these two areas and offer an alternate system that incorporates the strongest features of the systems reviewed.

THE PROBLEM OF UNIT OF ANALYSIS

Most commonly used systems for the analysis of written text use the sentence, t-unit, clause or proposition as the unit of analysis. By far the most commonly used unit in writing research is the t-unit. For example, the systems developed by Halliday and Hasan (1976), Nold and Davis (1980) use the t-unit. The system developed by Cooper and Matsuhashi (in press) were meant to be used with the sentence but are easily adapted to use with the t-unit. Such a large unit of analysis as the sentence or t-unit allows for rapid coding of large blocks of text but it ignores some rather salient features of discourse connectedness. To illustrate, consider the following t-unit taken from the Nold and Davis articles.

> ...but the amazement which War and Peace revives in me while I am reading it is like that of a child.

The relationship this t-unit has to the previous one is coded as conjunctive; this relationship is signaled by the conjunction but. However, the function of the clauses "which War and Peace revives in me" and "while I am reading it" are lost in the analysis. The t-unit

or sentence, then, appears to be too large a unit of analysis for a comprehensive research tool. Nold and Davis justify the use of the t-unit on the grounds that research into the reading process indicates that a reader keeps the elements of the t-unit in short term memory until the end of the t-unit is reached. This finding is common within reading research. However, it is not consistent with the research which indicates that the clause (both finite and nonfinite) is a salient unit of language processing (Bever, Garrett & Hurtig, 1973). It also ignores the writing research which indicates that t-unit length varies across age, grade and quality levels. This implies that a unit which recognizes all clausal structures and which accounts for variations in t-unit length might be a better unit of analysis. Such a unit is the proposition.

A proposition is defined as "two or more concepts, such that the first is a relation and the others arguments." (Turner & Greene, 1977, p.4). Propositions can be classified into three distinct groups: predication, modification and connective. To illustrate, consider the following:

The tall boy skated like a professional.

Two propositions are predication propositions:

(skate: A: boy)

(skate: A: professional)

One is a modification proposition:

(quality of: boy, tall)

One is a connective proposition:

[contrast: like, A_1(skate, boy), A_2(skate, professional]

This type of analysis is highly specific and captures the majority of explicit relationships within the text. In fact, propositional analysis captures perhaps the most basic type of relationship among units of thought - that among concepts - at what might be likened to a deep structure level.

Propositional analysis was developed as a methodology for the representation of meaning in long term memory. Although it accomplishes this task well it is perhaps too specific for writing research purposes. That is, the unit of analysis is too small. A thorough analysis using a propositional system identifies too many relationships to be an efficient language analysis tool. To illustrate consider the Turner and Greene analysis of the following paragraph into over twenty propositions:

> Louise and Ann went to the movies last night. They met Charlie there. Afterwards they all went for a chocolate sundae, but the ice cream parlor was closed.

1. (CONJ: AND, LOUISE, ANN)
2. (GO, A:1, G:MOVIES)
3. (QUALIFY, NIGHT, LAST)
4. (TIME: DURING, 2, 3)
5. (MEET, A:1, O:CHARLIE)
6. (LOC: AT, 5, MOVIES)
7. (CONJ: AND, LOUISE, ANN, CHARLIE)
8. (GO, A:7, S:MOVIES, G:12)
9. (QUALITY OF, SUNDAE, CHOCOLATE)
10. (GET, A:7, O:9)
11. (PURP: FOR, 8, 10)
12. (QUALIFY, PARLOR, ICE CREAM)
13. (CLOSE, A:$, O:12)
14. (NEGATE, 10)
15. (CAUS: BECAUSE, 13, 14)
16. (CONC: BUT, 11, 14)
17. (TIME: AFTERWARDS, 5, 11)
18. 0
19. 0
20. 0

17a. (END, A:MOVIES)
18a. (TIME: AFTERWARDS, 17a, 11)

19a. 0
20a. 0

17b. (END, A:MOVIES)
18b. (TIME: AFTERWARDS, 17a, 11)
19b. (TIME: AFTERWARDS, 11, 17b)
20b. (CONJ: ALSO, 18b, 19b)

If the sentence and t-unit are too large for effective analysis of written text and the proposition too small, the clause (both finite and nonfinite) might be a useful compromise unit. A system of analysis which considers only finite clauses would consider the following as a single unit:

> The boy laughing and playing near the swing watched the girl running near the stream.

A system which recognizes both finite and nonfinite clauses would divide this into five units:

(1) The boy watched the girl

(2) (boy) laughing

(3) (boy) playing near the swing

(4) (girl) running near the stream

This unit has been labeled the predication (Marzano, 1983). A predication contains "an explicitly stated verb called the predicate and one or more nouns, pronouns, adjectives, adverbs, interjections or wh-words called arguments. The purpose of the verb is to 'predicate' or establish information about the arguments." (p. 14)

The predication captures the fundamental relationship between verb and argument but does not describe that relationship as is done in propositional analysis. To accomplish this we must use a case grammar approach. That is, we must label the various types of relationships

that may exist among arguments and the predicate within a predication. In the system described here there are 25 argument or case types. They are exemplified but not described below.

Argument Type	Example
Agent	Bill kissed Mary.
Instrument	The key opened the door.
Stimulus	He was tired from the trip.
Affected	Bob's mind deteriorated.
Described	Jana is beautiful.
Recipient	Sylvia gave Carmen a cookie.
Subject Matter	They talked about football.
Direction	He went outside.
Object	I like candy.
State	He was sleepy.
Class	He is a politician.
Standard/ Comparison	He is quick for a large man.
Owner/Possessor	The car is mine.
Nonrelative Time	The dance was on Friday.
Duration	Bob talks incessantly.
Location	The car was parked near the fence.
Quality	He is tall.
Existence	It was raining.
Situation	He was on vacation.
Attitude	Frankly, I'm tired.
Exclamation	Boy! Was he good.
Intensifying/ Diminishing	He was nearly exhausted.
Manner	He ran quickly.
Accompanier	Mike along with Jack went to the store.
Specifier	There were four people in the car.

This extended case system is similar to that suggested by de Beaugrande (1980). In effect an expanded case system provides us with a descriptive ability at about the same level of specificity as propositional analysis. That is, we can describe the interrelationship among concepts within our unit of analysis. To illustrate consider the following:

> He left the tired, lonely man near the ancient bridge on the river.

If we were to propositionally analyze this predication we would obtain the following six propositions:

1) (left: A: he: O: man)

2) (quality of: man: tired)

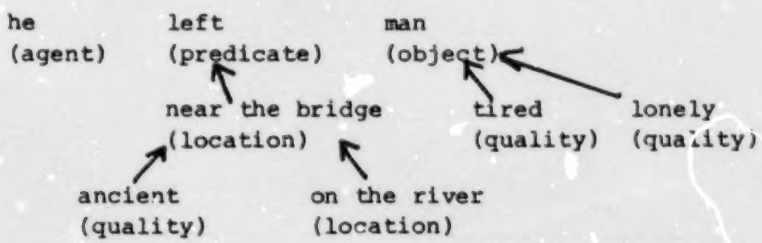
3) (quality of: man: lonely)

4) (location: near, 1, bridge)

5) (quality of: bridge: ancient)

6) (location: on, bridge: river)

If we were to use the predication as the unit of analysis and the argument categories described above we would obtain the following:



This type of argumental "mapping" conveys similar information to propositional analysis; argument types are identified along with an identification of sets of related arguments. Again this mapping procedure is similar to that suggested by de Beaugrande (1980).

RELATIONSHIPS BETWEEN PREDICATIONS

Given the predication as a unit of analysis, the next step in developing a "compromise" language analysis system is to identify the possible relationships that may exist among the units of analysis. As we have already seen the issue of unit of analysis is confounded with the issue of relationship types. When the unit of analysis is large (e.g. the t-unit) salient relationships between elements within the unit are lost. For example, given the t-unit as the unit of analysis relationships between embedded clauses are not coded. When the unit of analysis is small (e.g. the proposition), many more possible relationships must be accounted for.

In the system described here there are two basic types of relationships each with subtypes. Those are argumental and non-argumental relationships. An argumental relationship exists between two predications when they have an argument in common or when an argument in one predication refers to an entire predication or set of predications. For example, predications a and b below share the argument car:

a) Bill's car is beautiful.

b) It is the nicest on the block.

As the name implies, non-argumental relationships are those that do not require a shared argument. To illustrate consider the following two predications:

(Mark went home) because (the party ended.)

predication a predication b

Here there is no explicitly stated shared argument in the surface structure of the two predications (although we might argue that there

is an implied direction argument "from the party" in the first predication). However, there is an obvious causal relationship between the four predications. Causality, then, is a type of non-argumental relationship because its existence does not require a common argument between the two related predications. This does not imply that the existence of one type of relationship precludes the existence of the other. Indeed, most discourse is connected by overlapping networks of argumental and non-argumental relationships.

There is ample theoretical and intuitive support for separating predication relationships into these two broad categories. Argumental relationships are basically synonymous with the types of relationships called "reference" in other systems. Most of the literature on comprehension of connected discourse postulates that reference is the type of relationship first searched for in language processing. For example, Kintsch and van Dijk (1978, p. 367) state that the first step in processing a set of connected discourse (forming what is called a text base) is to check referential cohesion or the argumental overlap among propositions. If this overlap is found then all propositions are accepted for further processing. If an overlap is not found then inferences are made to "fill in" missing information. Argumental relationships also form the basis for the topic-comment and presupposition-assertion relationships described by Haviland and Clark (1974).

Argumental relationships, then, appear to be such a unique type of linkage between ideas that they warrant a category unto themselves. The various types of argumental relationships are briefly described below.

1) Same Word Reference: The most obvious and straightforward way of signaling a shared argument between predications is to use the same word to express the concept in both predications:

e.g. Bill is my friend.

Everyone likes Bill.

2) Pronoun Reference: The most commonly used method of signaling argumental relationships is to use a pronoun to represent a previously stated concept:

e.g. Bill is my friend.

Everyone likes him.

Here I am restricting the pronouns used in this category to the following: I, me, you, we, us, he, him, she, her, they, them, it, these, those, there, here. The use of all of these pronouns is fairly obvious except for the last two. There and here signal pronoun reference when they refer to a noun which names a specific place:

e.g. I went to Chicago last summer.

There I found true love.

3) Synonym Reference: Synonym reference occurs when an argument in one predication is expressed as a synonym in another:

e.g. Yesterday Bill bought a new car.

This automobile is the answer to his prayers.

4) Metaphoric Reference: Metaphoric reference is similar to synonym reference except that a direct synonym is not used to represent the shared argument. Instead, a word or phrase not normally considered a synonym is used:

e.g. Bill went to Seattle for the summer.

This haven for the weary is located in the northwest.

5) Wh-word Reference: There are many variations to wh-word reference. In general, it is signaled whenever any of the following wh words are used to refer to some previously stated argument: who, whom, which, that, what, where, how, why, whoever, whatever, however, wherever. The first four wh-words, who, whom, which and that, are commonly used as what traditional grammarians call relative pronouns. In such situations they usually refer to the immediately preceeding argument:

e.g. I know the man who is over there.

a) I know the man

b) (who) is over there.

This is the car that I want

a) This is the car

b) (that) I want

In some situations the wh-word can be omitted because its existence is understood.

e.g. This is the car (that) I want.

When the wh-words where, how, why and what are used the argument they refer to (called the "antecedent argument") is commonly, if not always, omitted.

e.g. I know (the place) where he stayed.

I know (the manner) how he escaped.

I know (the reason) why he left.

I know (the object) what I want.

The wh-words whoever, whomever, whatever, however and wherever are used when the antecedent argument is not only deleted but is unknown to the speaker or writer:

e.g. Bill will marry (the person) whomever he likes.

6) Deleted Subject Reference: This form of argumental relationships always occurs in the following constructions: 1) participles, 2) infinities, 3) gerunds, and 4) compound predicates. To illustrate, consider the following:

e.g. The boy running down the street likes the girl sitting on the porch.

Here there are three predications:

a) The boy likes the girl

b) (boy) running down the street

c) (girl) sitting on the porch

It is understood that boy is the subject of predication b and girl the subject of c. This is the case with all constructions listed above. They always contain a deleted but understood argument in the subject slot.

7) Related Argument Reference: Related argument reference occurs whenever an argument is used in one predication which is logically related to an argument stated in a preceding predication. To illustrate, consider the following:

a) Bill bought a new car.

b) The sun roof is one of its best features.

c) The tires cost $500.

d) The engine has 450 horsepower.

Here the arguments roof, tires and engine are all considered "related" to car stated in predication a. There are three types of related argument relationships:

a) Part to Whole: The example above is part to whole. Roof, tires and engine are "parts" of the whole--car.

b) Member of a set or class: Consider the following:

a) Animals are intelligent beings.

b) Dogs can understand many things.

c) Chimps are considered as smart as human infants.

d) Even rats can learn complex tricks.

Above dogs, chimps and rats are members of the general set or class called animals.

c) Ownership/Possession: When an argument is introduced which is owned or possessed by another previously stated argument, the relationship is considered to be that of related argument. Usually possessive pronouns or nouns in the genetive case are used to signal this type of argumental relationship:

e.g. I like Bill.

His sister is nice too.

8) Whole Predication Reference: This type of argumental relationship between predication occurs when an argument in one predication refers to a previously stated, entire predication:

e.g. a) Bill hit Mark.

b) This bothered Mary.

The word this in predication b refers to the entire event stated in predication a. Technically there is no shared concept between the two. With all other forms of argumental relationship the sharing of a

concept is assumed. I choose to classify whole predication reference as argumental because in terms of form it follows the same pattern as all argument relationships; specifically an argument in one predication refers to something in a previous predication. There are other more technical similarities between whole predication reference and the other types of argumental relationships. For a detailed discussion see Marzano (1983).

Commonly the words this, that and which are used to signal whole predication reference:

e.g. a) Bill hit Mark.

b) which bothered Mary.

A special type of whole predication reference is signaled by the words there and it. They do not refer specifically to a predication but to the situation in which the event expressed in the predication occurred.

Non-argumental relationships fall into five broad categories: 1) causality, 2) addition, 3) contrast, 4) time and 5) embedded relationships. Each are briefly described below:

1) Causal Relationships: Causal non-argumental relationships indicate that one event or state was the cause for another event or state. There are five sub-categories of causal non-argumental relationships:

a) Direct Cause:

He won the race by maintaining his concentration.

b) Result:

Bill went home. Consequently, the party ended.

c) Reason:

He went to the store because he needed food.

d) Inference:

Mary is going on a long trip. In that case she should plan well.

e) Condition:

Unless you stop, I will leave.

2) Addition Non-argumental Relationships: There are four sub-types of addition relationships. The essence of each is that the content of one predication is likened to the content of another predication on some dimension:

a) Equality:

He is tall, and he is handsome.

b) Restatement:

I am tired. In fact, I am exhausted.

c) Example:

He does many things well. For example, he is excellent at cards.

d) Summation:

He does many things well.

He cooks.

He sews.

In all he is an excellent homemaker.

3) Contrast Non-argumental Relationships: Where additive relationships emphasize the likeness of predications; contrast relationships highlight the difference. There are four sub-types of contrastive relationships:

a) Antithesis:

I will be there, but I won't be happy.

b) Alternative:

Either it will rain or it will snow.

c) Comparison:

Bill is tall. In comparison his brother is short.

d) Concession:

I don't like violence. Nonetheless, I'll meet you at the fights.

4) Temporal Non-argumental Relationships: Temporal non-argumental relationships signal the order of time in which the events stated in a set of predications occur. There are three sub-types of temporal relationships:

a) Subsequent Action:

They went to the game. Afterward they went to the dance.

b) Prior Action:

They went to the dance after they went to the game.

c) Concurrent Action:

Bill thought about Mary while Mary thought about Bill.

5) Embedded Non-argumental Relationships: To illustrate embedded non-argumental relationship consider the following:

That I play golf bothers my wife.

Here there are two predications:

a) (I play golf.)

b) _____ bothers my wife.

The function of the first predication is to act as an argument in the second predication. Whenever an entire predication acts as an argument in another predication we call the relationship embedded. We could

name the semantic function of the embedded predication. For example, the relationship a above has to b might be labeled stimulus. In general these relationships will parallel the case categories described previously.

Embedded predications commonly have the syntactic form of noun clause, gerund or infinitive. When the form is noun clause, the signal words "that" or "the fact that" are commonly used:

e.g. The fact that I play golf bothers my wife.

We should note here the semantic similarity between embedded non-argumental relationships and whole predication argumental reference. Consider the following:

a) Whole Predication Argumental Relationship

(Bill hit Mark.)

(This) bothered Mary.

b) Embedded Non-argumental Relationship

(That Bill hit Mark)

_____ bothered Mary.

As these examples illustrate many types of embedded non-argumental relationships are semantically identical to some argumental whole predications relationships. Indeed, in some discourse grammars these two types of relationships are classified under the same general heading. My reason for separating the two is based on form rather than function. As was stated previously, whole predication reference is signaled by a word (this, that, which) which has an argumental function within its own predication. With embedded relationships one predication is logically and syntactically incomplete without the other. These reasons seem strong enough to me to separate the two

types of relationships with the understanding and recognition of their semantic similarities.

COHESION VS. COHERENCE

As it relates to language analysis cohesion is generally defined as the ways ideas within discourse are linked together to create a sense of unity. Cohesion is a necessary but not sufficient condition for coherence. Coherence refers to the inter-relatedness of the "whole"; while there may be a sense of sections within a piece of writing being connected via cohesive devices, the sense of wholeness is accomplished by exophoric characteristics which relate the piece to a "real world" setting (Whitte and Faigley, 1982).

Measuring cohesion is a relatively easy task when compared with that of measuring coherence. Coherence by definition demands a reference to information beyond the text - information that is contained in one's schema, frame, script or plan relative to the information presented in the text (de Beaugrande, 1980).

One method of analyzing and quantifying cohesion is to identify the relative frequency of various types of lexical ties within discourse. This is the procedure used by Witte and Faigley. Using an outlier study of essays rated holistically for quality they found that high quality essays had more cohesive ties per t-unit and utilized lexical collocation as the predominant mode of connection. The problem with such an approach is that it provides only pairwise information about units of analysis. Cohesion is created by repeated, predictable use of lexical ties. At a very basic level cohesion is a function of the line by line or serial connectedness of discourse. According to Kintsch and van Dijk (1978) the reader stores only a few propositions

in short term memory buffer when processing text. It is the connectedness of these units that determine the cohesiveness of the discourse. This suggests that a system of cohesion analysis should be one in which the lexical tie linking one predication with the next is identified.

Identifying lexical reference from one predication to the next seems to be the logical method of analyzing cohesion especially in light of the evidence indicating that identification of referential ties is one of the first lines of discourse processing. Using a predication based system we might analyze lexical reference via the following process:

1) Identify the first argument with a referent in each predication.

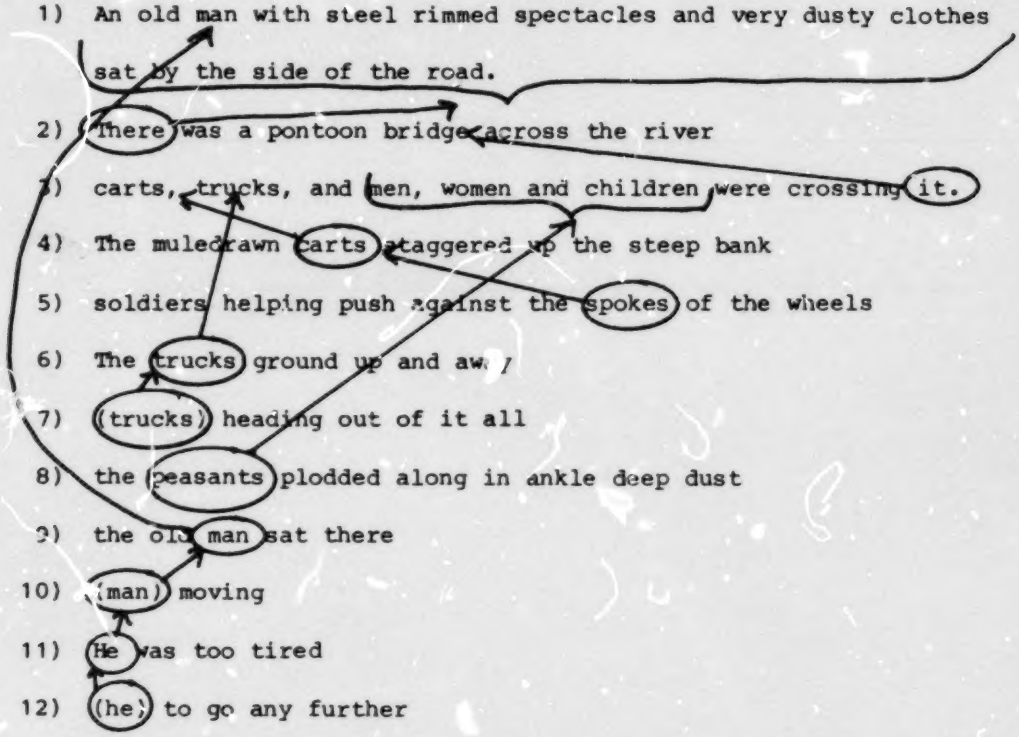
2) Connect that argument with its nearest referent.

I shall call this process mapping the primary line of reference. To illustrate consider the following passage:

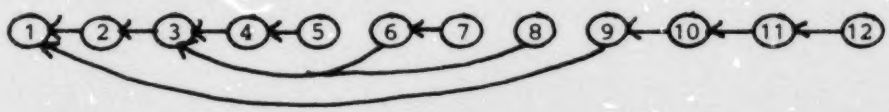
> An old man with steel rimmed spectacles and very dusty clothes sat by the side of the road. There was a pontoon bridge across the river, and carts, trucks, and men, women and children were crossing it. The muledrawn carts staggered up the steep bank from the bridge with soldiers helping push against the spokes of the wheels. The trucks ground up and away heading out of it all and the peasants plodded along in the ankle deep dust. But the old man sat there without moving. He was too tired to go any further.

Below I have mapped the primary line of reference:

1) An old man with steel rimmed spectacles and very dusty clothes sat by the side of the road.
2) (There) was a pontoon bridge across the river
3) carts, trucks, and men, women and children were crossing (it.)
4) The muledrawn (carts) staggered up the steep bank
5) soldiers helping push against the (spokes) of the wheels
6) The (trucks) ground up and aw /
7) (trucks) heading out of it all
8) the (peasants) plodded along in ankle deep dust
9) the old (man) sat there
10) (man) moving
11) (He) was too tired
12) (he) to go any further



This mapping gives us a good picture of the flow of argumental relationship from one predication to the next. If we represent each predication as a circled number and draw in the reference lines, the pattern becomes even clearer:



There are many ways of quantifying the extent to which predications are linearly connected within the primary line. One technique is to identify what I will call strings. A string consists of three or more predications which refer back to the immediately preceding predication. Strings can be identified by using the following procedure:

1) Identify the predication number of the primary line referent for each predication:

A Predication	B Referent of Primary Line
1	-
2	1
3	2
4	3
5	4
6	3
7	6
8	3
9	3
10	9
11	10
12	11

2) Subtract column A from column B.

A	B	B-A	
1	-	-	
2	1	1	
3	2	1	
4	3	1	string
5	4	1	
6	3	3	
7	6	1	
8	3	5	
9	3	6	
10	9	1	
11	10	1	string
12	11	1	

3) Analyze the B-A column for "runs" of 1. Above there is a string which runs from predication 2-5 and one which runs from predication 10-12.

There are a number of indices which can be calculated using strings. These include: a) length of longest string, b) average length of strings, c) range of string lengths. These indices can be used to compare the cohesive patterns of different samples of written discourse. Again, these and other indices which can be used to measure cohesion are described in Marzano (1983).

Coherence is a far more elusive entity than cohesion primarily because it is a function of characteristics outside of the text. Given that the purpose of this paper is to suggest methods of text analysis which might be used to quantify coherence, the exophoric nature of coherence indicates that any system which attempts to analyze explicit coherence signals will have inherent limitations. Before we can develop a system for coherence analysis we must develop a working or operational definition of coherence. Within models of long term memory storage of semantic information coherence is generally defined in terms of macrostructure.

The microstructure of text is the local level of the discourse, the structure of the individual propositions and their relations. The macrostructure is of a more global nature; it characterizes the discourse as a whole. A discourse is coherent only if its respective sentences are connected at the microstructure level and at the macrostructure level (Kintsch and van Dijk, p. 363). The macrostructure of text is determined via the application of macrorules. In brief, these rules are designed to: 1) identify those explicitly stated propositions which "govern" or subsume other propositions and 2) construct summary propositions where needed (van Dijk, 1980). We might conclude, then, that the process of determining coherence is not a line by line (predication by predication) phenomenon. Whereas cohesion is. Coherence is determined by applying a set of rules (macrorules) to the discourse as a whole. For a set of predications the macrostructure may be signaled by those predications which summarize or subsume the other predications in the set. A question germane to identifying macrostructure, then, is: "What type of predications or relationship

govern other predications?" Here I postulate that there are fine types of patterns of relationships each with governing or summarizing predications. I will call such structures macropatterns.

Topic Macropatterns

Topic macropatterns consist of a set of predications linked by argumental relationships and possessing the following characteristics: a) the predications state characteristics of an argument(s) within the primary line of reference and b) a single generalization introduces the set of characteristics or c) all arguments within the primary line of reference have a related concept relationship with a single stated argument. To illustrate consider the following contrived example:

a) Bill is nice.

b) He plays with his little sister.

c) He goes to church on Sunday.

d) He visits the sick.

Here the topic is "Bill". The pattern is governed by predication a - the macropredication of the set. Topic macropatterns also occur when a series of related argument references are used in the primary line of reference, particularly part to whole and set membership related argument reference.

a) There were many animals in the zoo.

b) The elephants ...

c) The lions ...

d) The birds ...

Here the argument "animals" governs the other arguments in the primary line of reference. Such a concept has been called a "super

concept" with regards to its position in the macrostructure (van Dijk, 1980).

Generalization Macropatterns

Generalization macropatterns are those in which predications are governed by a generalization but the predications within the set do not have a common argument within the primary line of reference.

a) Sometimes life gets tough.

b) Finances become a problem.

c) A period of poor health may develop.

d) Family problems can crop up.

e) Work may become dull and boring.

Here predication a (the macropredication) governs predication b-e; however, these predications do not share a common argument within the primary line.

Similarity/Dissimilarity Patterns

In such patterns the macropredication establishes the fact that one set of predications is different from or similar to another set.

a) The two fighters were very different.

b) The first one entered the ring with a great deal of bravado.

c) In contrast the other entered with little fanfare.

d) The first one wore a brightly colored robe.

e) The other wore black.

f) The first fighter used fancy footwork and quick jabs.

g) By comparison the quiet one protected himself and waited for an opening.

h) The first one scored many points on style.

i) The quiet one won by a knockout.

Here the macropredication (a) cues the reader that the following predication will have a dissimilarity pattern. Many times the macropredication within a similarity/dissimilarity pattern is not at the beginning of the set.

a) The first fighter was very flashy.

b) He came in with a great deal of bravado and

c) he wore a brightly colored robe.

d) He used fancy footwork and

e) scored many points on style.

f) The second fighter was equally as impressive.

g) He entered with a large entourage and

h) wore a black satin robe.

i) He relied on cunning and speed.

j) He, too, scored many points on style.

Here predications a and f can be considered macropredications. Predication a is a generalization identifying the dimension on which the fighters are alike. Predication f establishes the similarity relationship between set b-e and g-j.

Sequence Macropatterns

Sequence patterns are those in which time relationships are repeatedly used. Again a macropredication cues the reader that a sequence of events will be stated.

I have a regular routine in the morning.

First I ...

Then I ...

Next I ...

Finally I ...

Process Macropatterns

Process patterns are a set of ordered events which produce a specific result:

a) Baking a cake is easy.

b) First, get out the pots and pans.

c) Then, mix in the sugar, eggs and flour.

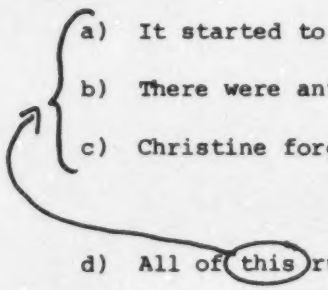
d) Then put the mix in the oven.

e) Twenty minutes later you have a cake.

Here the macropredication usually contains an embedded predication (e.g. "baking a cake") since most processes must be described by a verb form.

Embedded Set Macropatterns

We have already seen that a predication can be embedded within another predication and have an argumental function within the matrix predication. This same organization can occur at the macropattern level.



a) It started to rain.

b) There were ants.

c) Christine forgot the silverware.

d) All of this ruined the picnic.

Here the macropredication is predication d.

Patterns Without Macropredication

Four of the five macropatterns mentioned above can be stated without their macropredications. For instance we may have a topic pattern without the governing macropredication.

Bill plays with his little sister.

He goes to church on Sunday.

He visits the sick.

Similarly generalization, sequence and process patterns can be stated without their macropredications. By definition embedded set patterns contain a macropredication.

Patterns without macropredications will simply be called "patterns". When the reader encounters patterns he/she must construct a macropredication to store the information in long term memory.

Analyzing Discourse for Coherence

We may now operationally define coherence in discourse as the existence of macropatterns. In fact we may hypothesize that a set of discourse is most coherent if it contains macropatterns; less coherent if it contains explicit patterns and even less coherent if it contains no patterns or patterns which are hard to identify. The process of analyzing text for coherence, then, is that of identifying macropatterns and patterns. To illustrate consider the following:

> Fundamental laws and principals cannot be deduced. One knows them by creating them from nothing, out of one's Self. One does not arrive at fundamental laws and principles as a function of what is already known. Such laws and principles do not merely explain; they illuminate. They do not merely add to what we know; they create a new space in which knowing can occur.

The predicational structure of this passage is as follows:

1) Fundamental laws and principals cannot be deduced.

2) One knows them

by

3) (one) creating them from nothing, out of one's Self.

4) One does not arrive at fundamental laws and principles

as a function of

5) what is already known

6) Such laws and principles do not merely explain

7) they illuminate

8) They do not merely add to _____

9) what we know

10) They create a new space

11) in which knowing can occur.

Here there are two macropatterns - both generalization macropatterns. Predication 1 is a macropredication. Predications 2 and 4 are examples of the macropredication. Predications 4 and 5 considered together are a macropredication. Predications 7 and 10 are examples of that generalization.

We might quantify the coherence of this passage via a simple index which is formed by the ratio of the number of predications within a macropattern over the total number of predications. In this case that index would be 6/11 or .545. There are a number of ways of increasing the interpretability and, perhaps, even the validity of this index. One way might be to shift the unit of analysis from the predication to the t-unit. This makes intuitive and theoretical sense. Recall that from a language processing perspective the assumption is that the information processor shifts to a higher level of analysis when identifying macrostructures. It makes sense, then, that the unit of analysis would also shift. This position is corroborated by the fact that many generalizations extend beyond the predication level. If we

were to analyze the sample passage in terms of t-units we would obtain the following results.

1) Fundamental laws and principals cannot be deduced.

2) One knows them by creating them from nothing, out of one's Self.

3) One does not arrive at fundamental laws and principles as a function of what is already known.

4) Such laws and principles do not merely explain

5) they illuminate

6) They do not merely add to what we know

7) they create a new space in which knowing can occur.

Here the first generalization pattern extends from t-unit 1 (the generalization) through 3. T-unit 3 is the generalization for the second pattern. Units 5 and 7 are examples of the generalization. Forming the coherence ratio using the t-unit we obtain 5/7 or .714.

Here I should note that macropatterns and patterns are probably only the first level of what cognitive psychologists consider the macrostructure of discourse. Indeed, the elements of the macrostructure are probably themselves arranged in a hierarchical fashion the top level of which might be the title of the passage. Also there is at least another dimension of text coherence which might be called the superstructure (van Dijk, 1980). That dimension is not considered in this analysis.

A Cohesion/Coherence Display

The argumental relationships, non-argumental relationships and macropatterns that exist within a set of predications can be displayed

in a matrix that serves as a good visual summary for the cohesion/coherence ties within discourse. Such a cohesion/coherence display is illustrated below. Note that this display is hypothetical. It is contrived for illustrative purposes:

Predication #	Argumental Referent/type	Non-argumental Referent/type	Macropatterns		
1	-		A		
2	1: pronoun		X	B	
3	2: pronoun		X	X	
4	2: wh	3: cause		X	C
5	4: pronoun	4: time	X		X
6	4: related		X	X	X
7	6: whole pred.	6: contrast		X	
8	7: pronoun	7: addition			X
9	8: same word	8: time	X		
10	1: related		y		

A = generalization
B = topic
C = process

Above, the first column contains a simple listing of the predications. Here only the predication number has been listed. Column 1 might also contain a transcription of the actual predications. Column 2 contains the primary line, argumental referent for each predication and the type of argumental relationship. For example, predication 2 has predication 1 as its referent within the primary line; it has a pronoun argumental relationship with 1. Predication 6 has predication 4 as its referent within the primary line; it has a related argument relationship with 4. The third column contains the non-argumental referent for each predication and the type of relationship. For example, predication 5 has a time non-argument relationship with 4; predication 8 has an addition non-argumental relationship with 7. The last three columns are under the general heading of macropatterns. It is in this section of the display that

all the macropatterns within a set of predications are identified. The first macropattern (labeled A) is a generalization macropattern which includes predications 1, 2, 3, 5, 6, 9 and 10 (an x indicates that a predication is part of the pattern). The second macropattern (labeled B) is a topic pattern which includes predications 2, 3, 4, 6 and 7. The third macropattern, C, is a process pattern which includes predications 4, 5, 6 and 8.

Such a cohesion/coherence display might be more than just a data organization technique. Indeed, it might identify the various levels or layers of processing involved in the comprehension of connected discourse. Argumental and non-argumental relationships are already a component of virtually all text comprehension models. A model which attempts to explain all top-down comprehension processes must also include slots for macropattern and even higher level organizational information within the short term memory buffer system (Kintsch, 1979; van Dijk, 1980).

SUMMARY

A unit of language analysis called the predication was defined and described. Such a unit is somewhat of a compromise unit between the proposition and the t-unit or sentence. Argumental and non-argumental relationships between predications were identified. Cohesion was operationally defined as the predication by predication connectedness of the discourse. Coherence was defined as the existence of explicitly stated patterns (macropatterns) which are part of the macrostructure of the discourse. Quantitative techniques were suggested for both cohesion and coherence analysis.

REFERENCES

Bever, T., Garrett, M. and Hurtig, R. "The Interaction of Perceptual Processes and Ambiguous Sentences." Memory and Cognition, (1973), Vol. 1, pp. 277-286.

Cooper, C. and Matsuhashi, A. "A Theory of the Writing Process" in Marllen (Ed.) The Psychology of Writing. London: Wiley. In press.

de Beaugrande, R. Text, Discourse and Process. Norwood, N.J.: Ablex, 1980.

Halliday, M. and Hasan, R. Cohesion in English. London: Longman, 1976.

Haveland, S. and Clark, H. "What's New? Acquiring New Information as a Process in Comprehension" Journal of Verbal Learning and Verbal Behavior. (1974), Vol. 13, pp. 512-521.

Kintsch, W. "On Modeling Comprehension" Educational Psychologist, (1979), Vol. 14, pp. 3-14.

Kintsch, W. and van Dijk, T. "Toward a Model of Text Comprehension and Production" Psychological Review, (September, 1978). Vol. 89, pp. 303-394.

Marzano, R. A Quantitative Grammar of Meaning and Structure. Denver: Mid-continent Regional Educational Laboratory. 1983.

Nold, E. and Davis, B. "The Discourse Matrix," College Composition and Communication (May, 1980), Vol. 31, pp. 141-152.

Turner, A. and Greene, E. The Construction and Use of a Propositional Text Base. Boulder, CO: The Institute for the Study of Intellectual Behavior, 1977.

van Dijk, T. Macrostructures. Hillsdale, NJ: Leaurence Erlbaum Associates, 1980.

Whitte, S. and Faigley, L. "Coherence, Cohesion and Writing Quality." College Composition and Communication. (May, 1982), Vol. 32, pp. 189-204.